Geoderma 209–210 (2013) 65–74

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Land use and lead content in the soils of London

R.M. Lark *, C. Scheib

British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form 4 June 2013
Accepted 9 June 2013
Available online 4 July 2013

Keywords:
Lead content
Urban soils
Land use
Orthogonal contrasts
Cokriging

ABSTRACT

It is important to understand how and where pollution and other anthropogenic processes compromise the ability of urban soil to serve as a component of the natural infrastructure. An extensive survey of the topsoil of the Greater London Area (GLA) in the United Kingdom has recently been completed by a non-probability systematic sampling scheme. We studied data on lead content from this survey. We examined an overall hypothesis that land use, as recorded at the time of sampling, is an important source of the variation of soil lead content, and we examined specific orthogonal contrasts to test particular hypotheses about land use effects. The assumption that the residuals from land use effects are independent random variables cannot be sustained because of the non-probability sampling. For this reason model-based analyses were used to test the hypotheses. One particular contrast, between the lead content in the soil of domestic gardens and that in the soil under parkland or recreational land, was modelled as a spatially dependent random variable, predicted optimally by cokriging.

We found that land use is an important source of variation in lead content of topsoil. Industrial sites had the largest mean lead content, followed by domestic gardens. Detailed contrasts between land uses are reported. For example, the lead content in soil of parkland did not differ significantly from that of recreational land, but the soil in these two land uses, considered together, had significantly less lead than did the soil of domestic gardens. Local cokriging predictions of this contrast varied substantially, and were larger in outer parts of the GLA, particularly in the south west.

1. Introduction

Traditionally soil survey and inventory has been focussed in the rural environment, to provide information for agricultural extension, catchment management, etc. In the last 25 years or so there has been increasing interest in the soil of urban environments (Bullock and Gregory, 1991). It is recognized that these soils are influenced by human activity to a unique extent (Craul, 1985) and that urban soil quality can have a direct influence on human health. In particular humans are exposed to contaminants present in the soil both through inhalation and ingestion (Mielke et al, 2007), and this may have direct consequences for health and wellbeing (Miranda et al., 2007).

This growing awareness of the importance of urban soils has resulted in increasing activity to survey them. Johnson et al (2011) provide accounts of geochemical surveys of urban soils in several European cities, and urban sites have been introduced into the network of long-term ecological research sites in the United States (Bain et al., 2012). In 1992 the British Geological Survey (BGS) extended its geochemical baseline survey of the environment, which includes soil sampling, to include urban soils (Fordyce et al., 2005). Since then more than 20 major urban centres in the UK have been surveyed. The resulting data have been used to study the spatial distribution of potentially hazardous elements in soil, notably metals (e.g. Marchant et al., 2011; Rawlins et al., 2005).

Lead in urban soil is of considerable interest. It is strongly influenced by human activity, and diffuse lead pollution of the soil arises from industrial use of lead and from the past use of lead tetraethyl as an additive in petrol. In the past lead piping was extensively used for plumbing and lead was a major constituent of paints. All these anthropogenic sources of lead are particularly dense in the urban environment, and so lead is commonly elevated both in urban soils (Clark et al, 2006) and in periurban soils (Rawlins et al, 2012). Furthermore, lead is persistent in soils and sediments because it is strongly bound by various soil minerals (Maurice, 2009), so historical land use may have a pronounced 'signature' in the contemporary content of soil lead. This lead is a threat to human health. The resuspension of soil lead is an important source of atmospheric lead, which may be inhaled and so absorbed (Laidlaw and Filippelli, 2008). The ingestion of soil is another important pathway by which soil lead can be absorbed, particularly by children (Guney et al., 2010). Another pathway for the uptake of lead is possible if some urban soils, such as domestic gardens, are used to grow vegetables for human consumption (Davies et al, 1979).

It is therefore important to understand the origins and distribution of lead, as well as other metals, in the urban soil. In the present paper we use data from a recent extensive survey of the topsoil across the whole of London to examine specific orthogonal hypotheses about

* Corresponding author.
E-mail address: mlark@nerc.ac.uk (R.M. Lark).

0016-7061 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY license.
http://dx.doi.org/10.1016/j.geoderma.2013.06.004

how land use, recorded at the time of sampling, accounts for variations in total soil lead content. This requires a model-based analysis because the survey, like many others carried out to provide geochemical baseline data, was not conducted by probability sampling. Such an analysis provides estimates of mean effects (such as the difference in lead content between two particular land uses) but these effects may vary spatially in ways that give insight into the nature of the land use effect. We therefore examine how a particular land use effect on soil lead content varies spatially across the Greater London Area using an optimal cokriging estimator.

2. Materials and Methods

2.1. The London Earth survey

From 2005 to 2009 BGS undertook soil sampling in the Greater London Authority (GLA) area which comprises the City of London and the London Boroughs (Knights and Scheib, 2010). Details of the survey methodology are provided by Johnson (2005) but we summarize the key points here.

Sampling was undertaken according to a non-probability systematic design. Each 1 × 1-km square of the British National Grid within the GLA area was sampled, with a sample site in each of the four 500 × 500-m quadrants. The sample site was placed as close as possible to the centre of the quadrant avoiding any obvious sources of point contamination such as spoil heaps. The surveying team identified the land use from a series of land use codes (see Table 1) at each sample site. At each site a soil specimen was collected with an auger at the centre and vertices of a 20 × 20-m square centred on the site. The soil specimen was collected with an auger from depth 0 to 15 cm after removal of any surface litter. The five specimens were then aggregated to form a single bulk specimen for the site. At one sample site in every 100 a duplicate aggregate specimen was formed from cores collected at the centre and vertices of an adjacent 20 × 20-m square.

The aggregated material from each site was subsequently air-dried, disaggregated and sieved through a nylon sieve to pass 2 mm and sub-sampled by coning and quartering. A 50-g sub-sample was ground in an agate planetary ball mill until 95% of the material was finer than 53 μm. Concentrations (totals) of 50 major and trace elements were determined for each sample by wavelength dispersive X-ray fluorescence (XRF) spectrometry. Data were obtained for a total of 6245 sites.

2.2. Exploratory data analysis

Summary statistics were computed for the data on lead content and a transformation to logarithms was considered. The residuals of lead content from land use class mean were also examined, both with and without transformation of the original data. In general if the conventional coefficient of skewness is out with the range [−1.0, 1.0] then a transformation is deemed necessary (Webster and Oliver, 2007). However, the conventional coefficient of skewness is very susceptible to outlying values, and decisions on transformation should be based on the shape of the distribution of the underlying variable, separate from the effects of any outlying values. For this reason we used the octile skewness, which is a robust measure of the asymmetry of the distribution of a variable, insensitive to outliers (Brys et al., 2003). An equivalent rule of thumb to that of Webster and Oliver (2007) for interpretation of the conventional skewness coefficient is to consider transformations of a variable if the octile skewness is not in the range [−0.2, 0.2] (Lark et al., 2006). By basing decisions on transformation on the octile skewness we aim to use transformations only when this is necessary to justify the assumption that the data are drawn from an underlying normal random variable, possibly with outliers present. Note that, because some of the recorded lead concentrations were zero it was necessary to add a small constant (0.1) to the data before the transformation.

Table 1
Land use classes and numbers of observations in each.

Class	Number of observations
Arable	270
Commercial and residential	195
Domestic garden	1554
Industrial	64
Other	197
Park	828
Pasture	144
Recreational	657
Road verge	597
Rough grazing	346
Urban open space	1119
Woodland or forest	274

2.3. Overall land use effects

2.3.1. The model, estimation and inference

Table 1 shows the land use classes that were considered in this study. The objective was to examine the evidence for overall differences between the classes with respect to lead content, and to consider more specific questions about contrasts between particular land uses or groups of land uses.

One method to address such a question is the analysis of variance, with the partition of the sum-of-squares for differences between p land use classes into components that correspond to particular contrasts. That method cannot be used in this case because the analysis of variance is based on the assumption that the residuals from land use class means can be regarded as independent random variables. That assumption is justified by the use of an appropriate probability sampling design, such as simple random sampling or stratified random sampling (de Gruijter et al., 2006). Such an assumption cannot be made with these data because the sample sites are selected according to a systematic rule with no element of randomization. For this reason a model-based analysis is necessary. We regard the $n \times 1$ vector of observations $\mathbf{z}$ as a realization of a random function, $\mathbf{Z}$ which is described by the linear mixed model

$$\mathbf{Z} = \mathbf{M}^T\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{M}$ is a $n \times p$ vector that indicates the land use class present at each location so that the element in the ith column and jth row is 1.0 if the ith class occurs at the jth sample location, and is zero otherwise. The $p \times 1$ vector $\boldsymbol{\beta}$ contains the fixed effects coefficients which here are mean values of the variable of interest within each land class. The terms $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are, respectively, a spatially correlated and an independently and identically distributed random variable. Ideally the components of a linear mixed model are estimated by finding the residual maximum likelihood (REML) estimates of the variance parameters of the random variables, and then generalized least squares estimates of the fixed effects coefficients (e.g. Lark and Cullis, 2004). However, this is not practical with substantial data sets such as this one, with $n =$ 6245, since the REML estimation requires repeated inversion of a $n \times n$ matrix, which is computationally demanding. It was therefore decided to estimate the parameters of the random components from ordinary least squares (OLS) residuals from the class means. REML is generally preferred because estimates based on OLS residuals are prone to bias because of estimation error of the class means. However, in a large data set this estimation error, and so the resulting bias, will be small.

For this reason we used an iterative generalized least squares procedure to fit the model (Webster and Oliver, 2007). The procedure is outlined below.

1. We found estimates of the land use class means by ordinary least squares (OLS) and computed the residuals. If $\hat{\boldsymbol{\beta}}$ denotes the OLS estimates of the land use-class means then the vector $\mathbf{y}$ contains the residuals

$$\mathbf{y} = \mathbf{z} - \mathbf{M}^T\hat{\boldsymbol{\beta}}. \tag{2}$$

2. The variogram of the residual component was estimated from the OLS residuals. We considered the standard estimator due to Matheron (1962). Although a log-transformation might be selected to deal with skewness of the underlying variable (Section 2.2) outliers may still be present, so the robust estimators due to Cressie and Hawkins (1980) and Dowd (1984) were also considered. Models were fitted to each of the resulting sets of estimates by weighted least squares implemented in the FVARIOGRAM procedure in GenStat (Payne, 2010), one model was selected for each set of estimates on the basis of the Akaike Information Criterion (AIC) (Webster and McBratney, 1989) and then the selected model for each of the sets of estimates was crossvalidated with the XVOK2D program in the GSLIB library (Deutsch and Journel, 1992). A model was selected from among those derived from the different estimators by computing from each cross-validation the standardized squared prediction error:

$$\theta(\mathbf{x}) = \frac{\left(\widetilde{Y}(\mathbf{x}) - y(\mathbf{x})\right)^2}{\sigma_K^2(\mathbf{x})} \tag{3}$$

where $\widetilde{Y}(\mathbf{x})$ denotes the ordinary kriging estimate of the residual at location $\mathbf{x}$, $\mathbf{y}(\mathbf{x})$ is the actual residual and $\sigma_K^2(\mathbf{x})$ is the ordinary kriging variance. We followed Lark (2000) in selecting the model for which the median value of $\theta(\mathbf{x})$ over all observations is closest to the expected value of 0.455 for normally distributed kriging errors with a valid variogram model.

3. The selected variogram model was then used to compute elements of the covariance matrix of the combined random components in Eq. (1), $\eta + \varepsilon$. We denote this matrix by $\mathbf{C}$, and obtain any element by

$$\mathbf{C}(i,j) = \gamma(\infty) - \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right), \tag{4}$$

where $\gamma(h)$ denotes the value of the selected variogram model for distance h, note that by definition $\gamma(0) = 0$.

4. The generalized least squares (GLS) estimate of the fixed effects coefficients in β were then obtained by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{M}^T\mathbf{C}^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^T\mathbf{C}^{-1}\mathbf{y} \tag{5}$$

The covariance matrix of the errors in the estimated class means in $\hat{\beta}$ is obtained by

$$\hat{\mathbf{V}} = \left(\mathbf{M}^T\mathbf{C}^{-1}\mathbf{M}\right)^{-1} \tag{6}$$

These expressions can be found in standard texts on linear modelling such as Dobson (1990).

5. We then return to step 1 of the procedure, but this time using the GLS estimate of $\hat{\beta}$, obtained with Eq. (5), to compute the residuals with Eq. (2). These steps are reiterated until the residuals show no further change.

Once this procedure has converged the resulting estimates $\hat{\beta}$ and $\hat{\mathrm{V}}$ provide estimates of the mean value within each land use class, and the variance of that mean respectively. It should be noted that these estimated means are model means, i.e. mean values for each land use class within the specified model (de Gruijter et al., 2006). Direct estimates of the spatial mean (i.e. the integral of lead content across all sites within a particular class) could be obtained by appropriate probability sampling, or by a model-based prediction if the extent of any particular land use class in the classification used here were known across the area. In the absence of such information the model mean is an appropriate alternative on which to base inferences about land use effects on this particular variable.

It is possible to examine a particular contrast, or a group of contrasts, between land use classes by computing an appropriate Wald statistic (see Lark and Cullis, 2004). For example, to test the null hypothesis that the first and second classes have identical mean values one proposes a $1 \times p$ contrast matrix, $\mathbf{L}_{1vs2}$:

$$\mathbf{L}_{1\ \mathrm{vs}\ 2} = [1 - 1 \quad 0 \ \ldots 0] \tag{7}$$

To test a null hypothesis that all classes have the same mean one would propose a $(p-1) \times p$ contrast matrix:

$$\mathbf{L}_{\mathrm{all}} = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ . & . & . & \ldots & . \\ 1 & 0 & 0 & \ldots & -1 \end{bmatrix} \tag{8}$$

Two sets of contrast $\mathbf{L_1}$ and $\mathbf{L}_2$ are orthogonal (i.e. independent) if and only if

$$\mathbf{L}_1\mathbf{L}_2^{\mathrm{T}} = 0. \tag{9}$$

The Wald statistic for some contrast matrix $\mathbf{L}$ is computed as

$$W = \left(\mathbf{L}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\left(\mathbf{L}\hat{\mathbf{V}}\mathbf{L}^{\mathrm{T}}\right)\left(\mathbf{L}\hat{\boldsymbol{\beta}}\right). \tag{10}$$

Under the null hypothesis W is distributed as χ^2 with degrees of freedom equal to the rank of $\mathbf{L}$, which is 1 for a single contrast matrix such as $\mathbf{L}_{1\ \mathrm{vs}\ 2}$ above and would be p-1 in the case of $\mathbf{L}_{\mathrm{all}}$.

2.3.2. Contrasts

The procedure for inference about contrasts outlined above is appropriate for contrasts identified prior to data analysis to test hypotheses of scientific interest. It should not be used post-hoc to test contrasts suggested by the data. We present below a set of contrasts identified prior to data analysis, with the associated rationale. The first test is an overall assessment of differences between land uses as a source of variation of lead content in the soil. In the second test we consider the overall effect of industrial land use. In contrasts 3–5 we consider some questions of interest concerning soils under land uses characteristic of the urban environment. Finally (contrasts 6–9) we consider how far soils under agricultural land use and woodland within the GLA contrast with soils under characteristically urban land uses, and examine the effects of intensity of management on the susceptibility of the soil to pollution.

1. *Differences among all class means.* First we test the initial (and recognizably implausible) null hypothesis that the mean lead concentration is the same under all land use classes in London.
2. *Difference between soil in land use class Industrial and all other land uses.* It is expected that industrial sites, in particular those with a long legacy of industrial use, may have been more subject to point contamination with lead than those under other land uses, but given the importance of atmospheric deposition of lead, and the potential for distribution over significant distances (McGrath and Loveland, 1992) it is of interest to assess how far industrial activity raises concentrations above the level of background contamination from diffuse sources, which is likely to be large in the urban environment.
3. *Difference between soil in land use class Parks and Recreational.* These two land uses constitute the principal areas of 'urban green space' under the classification, excepting woodland which we consider below. The ecosystem services provided by urban green space in the United Kingdom was recently assessed by the UK National

Ecosystem Assessment (2011) concluded that the annual economic value of urban green space was of the order of £30 billion, but also recognized that the quality of urban green space is under threat. In this contrast we assess whether these two categories of urban green space appear to differ with respect to lead content. This is possible since *Parks*, including the large Royal Parks in central and south-west London generally are larger, less fragmented areas than *Recreational* land.

4. *Difference between Parks and Recreational (considered together) and Domestic Gardens.* It is known that domestic gardens have historically been susceptible to contamination with lead due to activities such as the spreading of soot and coal ash on gardens, disposal of household waste on garden bonfires and contamination by flakes of lead-based paint (Thornton, 1991). In this contrast we examine whether these potentially large point sources of lead contamination impose a heavier lead burden on garden soils than is found in the soil of Parks and Recreational ground.
5. *Difference between Road verge and Urban open space.* It may be expected that soil of road verges is most directly exposed to pollution due to traffic, and there is some evidence for this (Chen et al., 2010). Because much lead pollution attributable to traffic has been due to emissions from engines using leaded petrol, it is possible that short-range transport to the road verge is relatively unimportant and these soils will not show significantly morelead than other soils in the urban environment.
6. *Difference between agricultural land (Arable, Pasture and Rough Grazing) and all other non-industrial urban land uses other than Industrial and Woodland or Forest.* Although the GLA is overwhelmingly urban, agricultural land is found within it. In general agricultural land might be expected to be less susceptible than other land uses to pollution by lead because it is generally distributed around the periphery of the GLA away from the most intense sources of lead pollution, but it is still susceptible to atmospheric deposition. Note that land use *Industrial* was not considered so that the contrast is orthogonal to contrast (2) above. Similarly *Woodland or Forest* was excluded so that a specific contrast, orthogonal to this one, could be examined to compare *Woodland or Forest* alone with agricultural land.
7. *Difference between Woodland or Forest and agricultural land (Arable, Pasture and Rough Grazing).* A priori one might expect that agricultural land is more susceptible than *Woodland or Forest* to atmospheric deposition of lead since some of this is intercepted in the latter case by the overlying tree canopy. However, some of this intercepted lead may be deposited on the soil due to leaf fall or leaching of the canopy by rain. It is of interest to assess contamination of woodland and forest soil in the urban environment. About 11% of the area of urban environments in the United Kingdom is woodland, and the particular value of urban woodland environments for human health and well-being was emphasized in the recent United Kingdom National Ecosystem Assessment (Davies, 2011).
8. *Difference between Arable and land under grazing (Pasture and Rough Grazing considered together).* Here we consider whether arable land differs from land under grazing with respect to lead content. Differences might be expected due to more intensive traffic on arable land, reduced cover during certain times of the year and historical application of wastes including sewage sludges.
9. *Difference between Pasture and Rough Grazing.* Here we examine whether land under *Rough Grazing* differs from more intensively-managed land under *Pasture* with respect to lead content of the soil.

2.4. Spatial variation of a land use contrast

The analyses described in section 2.3 above allow inferences to be made about the mean lead concentration in soils of different land use classes under the statistical model. However, it is possible that there is substantial spatial variation in any particular contrast across the GLA area reflecting some interaction of the land use effect with another spatial variable. For example, the magnitude of a contrast that reflects differences in land management, such as past application of wastes to soil, might vary with soil properties which influence the mobilization and leaching of applied contaminants in the soil (e.g. soil pH, soil texture). Bishop and Lark (2007) considered a similar case in agronomic field experiments in which the assumption of constant effects across variable regions may be implausible. They proposed a solution in which the effect of a treatment is not regarded as a fixed effect but rather as a random variable, for which point estimates or local regional means can then be estimated by a kriging predictor. It is that approach that we now consider.

In particular we consider the spatial variation of the fourth-listed contrast in section 2.3.2. This is the contrast between *Parks* and *Recreational* (considered together) and *Domestic Gardens*. The decision to examine the spatial variation of this contrast was made after completion of the analysis to test specific contrasts between the model means as described in section 2.3. This analysis showed that the null hypothesis of no difference between *Parks* and *Recreational* could not be rejected (contrast 3) but that the null hypothesis of no difference between *Parks* and *Recreational* (considered together) and *Domestic Gardens* could be rejected, with evidence for larger lead concentrations in the soil of domestic gardens (see Table 3). One may then consider whether the better soil quality (with respect to lead content) in *Parks* and *Recreational* than in *Domestic Gardens* is found everywhere in the GLA. This is important when weighing the policy significance of this land use effect. Is the relatively better soil quality in *Parks* and *Recreational* found in central London as well as in the suburbs, for example?

Bishop and Lark (2007) report on the analysis of an agronomic experiment, designed to answer the specific question about variation in treatment effects across an agricultural landscape. In the present study the spatial distribution of the land uses across the GLA is not under experimental control. Since the analysis that we use is model-based, rather than making any assumptions based on the sample design, this is not a fundamental problem. However, Bishop and Lark (2007) could ensure that the treatments of interest were distributed more or less uniformly across the study area so that there were no regions where the uncertainty of local estimate of a contrast between treatments would be large because one or more treatments of interest is underrepresented. However, as seen in Fig. 1, the distribution of *Parks* and *Recreational* (considered together) and *Domestic Gardens* as land uses across the GLA is reasonably uniform, although there are some areas of central East London where domestic gardens are sparse.

We now regard the lead content (after transformation to logarithms) at location x, where the land use is *Parks* or *Recreational* as a random function of location, $Y_{\mathrm{PR}}(\mathrm{x})$. The corresponding random function for land use *Domestic Gardens* is denoted $Y_{\mathrm{DG}}(\mathrm{x})$. In order to form a geostatistical prediction of the difference between these two random functions we require a linear model of coregionalization (LMCR) (Journel and Huijbregts, 1978) to describe their spatial variation and covariation. Since, by definition, the land use at any one location, x, cannot belong to both sets of categories there are no collocated data on both random variables, so the usual way to estimate a LMCR based on estimates of the auto- and cross-variogram functions (e.g. Webster and Oliver, 2007) cannot be followed. Papritz and Flühler (1994) proposed that the pseudo-cross variogram is used in these circumstances, and obtained an optimal cokriging estimator to predict the contrast between the two variables at any location. Papritz and Flühler (1994) suggested this approach for data from soil monitoring where the two variables of interest are measurements of the same soil property at different times. However, Bishop and Lark (2007) used the approach to estimate contrasts between responses to different experimental treatments, and the procedure can also be used for our problem of estimating the contrast between values of a soil variable under two land uses. We also followed Bishop and Lark

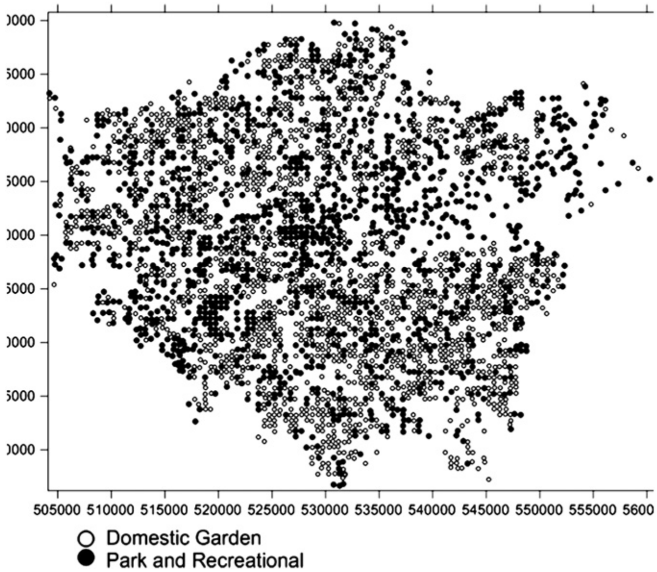

Fig. 1. Sample points allocated to land use classes *Domestic Garden* and *Park* or *Recreational* considered together.

(2007) by using robust estimators of the pseudo-cross variogram proposed by Lark (2002). The procedure used in this study is set out below.

1. Data on log-transformed lead content in the soil of the ith sample location in land use *Domestic Garden* and the jth sample location in land use *Park* or *Recreational* are denoted by $y(\mathrm{x}_i^{\mathrm{PR}})$ and $y(\mathrm{x}_j^{DG})$ respectively. Exploratory analysis showed no evidence of anisotropy (directional dependence) in these data, and so isotropic variograms were estimated.
2. Auto-variograms for the two land uses and the pseudo cross-variogram were estimated using the standard method of moments estimator (MoM) (Papritz et al., 1993). As explained in Section 2.3, any log-transformation of data is undertaken to deal with skewness of the underlying distribution of a variable, and outliers may remain, so two robust alternative estimators were considered, all of which are described in full by Lark (2002). The robust estimators are the square-root difference (SRD) estimator, which is equivalent to Cressie and Hawkins's (1980) robust auto-variogram estimator; and the median absolute difference (MAD) estimator, which is equivalent to Dowd's (1984) robust auto-variogram estimator.
3. Authorized variogram model (Webster and Oliver, 2007) was fitted to each set of auto-variogram estimates, by weighted least squares using the FVARIOGRAM procedure in GenStat (Payne, 2010), and a model was selected for each set of estimates on the basis of the AIC. Each model was then cross-validated by ordinary kriging, using the XVOK2D program in the GSLIB library (Deutsch and Journel, 1992) and the standardized squared prediction error, Eq. (3), was computed in each case. One estimator — MoM, SRD or MAD — was then selected after examining the median value of the standardized squared prediction error from each cross-validation.
4. An LMCR was then fitted to the selected set of estimates of the auto- and pseudo cross-variograms. This was done by weighted least squares, following the procedure set out by Lark (2002). The LMCR comprises a set of coregionalization matrices which are covariance matrices for modelled components of the set of variables. Each matrix has an associated authorized spatial correlation function specified by one or more distance parameters. From the LMCR it is therefore possible to compute the covariance or cross-covariance for the random functions which are modeled in the LMCR.
5. Papritz and Flühler (1994) propose a cokriging predictor for the difference between two random functions which is a special case of the general cokriging predictor of a linear combination of two random functions (Myers, 1983). In this case we formed block cokriging estimates of the difference between lead content in soil of *Domestic Gardens* and that of land in *Parks* or *Recreational* for 250 × 250-m square blocks. Block kriging was used for two reasons. First, because most management or regulatory decisions (on suitable land use or remediation) are made for blocks of land. Second, the nugget variances in the LMCR are relatively large and, as Burgess and Webster (1980) observed, block kriging for mapping emphasizes the dominant pattern of variation over short-range discontinuities. The estimate for a block denoted by $\mathcal{B}_0$ is given by:

$$\{Y_{\mathrm{DG}} \simeq Y_{\mathrm{PR}}\}(\mathcal{B}_0) = \sum_{i=1}^{u} \lambda_i^{\mathrm{DG}} y\left(x_i^{\mathrm{DG}}\right) - \sum_{j=1}^{v} \lambda_j^{\mathrm{PR}} y\left(x_j^{\mathrm{PR}}\right), \tag{11}$$

where x_i and $\mathrm{x_j}$ denote the locations of, respectively, the ith out of u neighbouring observations in land use class *Domestic Garden* and the jth out of v neighbouring observations in land use classes *Parks* or *Recreational*; λ_i^{DG} and λ_j^{PR} denote the corresponding cokriging weights. These are found by solving the cokriging equation:

$$\mathbf{K}\lambda = \mathrm{b}, \tag{12}$$

where

$$\mathbf{K} = \begin{bmatrix} \mathbf{C}_{DG}^{u,u} & -\mathbf{C}^{u,v} & 1^{u,1} & 0^{u,1} \\ -\mathbf{C}^{v,u} & \mathbf{C}_{PR}^{v,u} & 0^{v,1} & 1^{v,1} \\ 1^{1,u} & 0^{1,v} & 0 & 0 \\ 0^{1,u} & 1^{1,v} & 0 & 0 \end{bmatrix}. \tag{13}$$

Here $\mathrm{C}_{\mathrm{DG}}^{u,u}$ is a $u \times u$ matrix of the covariances of the lead content at the u neighbouring observations in domestic gardens. These covariances are obtained from the LMCR, as are the corresponding covariances for the v observations in land use *Parks* or *Recreational* which are in $\mathbf{C}_{\mathrm{PR}}^{v,v}$. The matrices $\mathbf{C}^{u,v}$ and $\mathbf{C}^{v,u}$ contain the cross-covariances of the neighbouring observations in the two classes; $1^{u,1}$ denotes a $u \times 1$ matrix with elements all equal to 1, and the remaining matrices are denoted by the same rule. The vector $\boldsymbol{\lambda}$ in Eq. (15) is defined as:

$$\lambda^{\mathrm{T}} = \left[\lambda_1^{\mathrm{DG}}, \lambda_2^{\mathrm{DG}}, \ldots, \lambda_u^{\mathrm{DG}}, \lambda_1^{\mathrm{PR}}, \lambda_2^{\mathrm{PR}}, \ldots, \lambda_v^{\mathrm{PR}}, -\psi_1, -\psi_2\right], \tag{14}$$

where ψ_1 and ψ_2 are Lagrange multipliers. The vector $\mathbf{b}$ in Eq. (15) is defined as:

$$\begin{aligned} \mathrm{b}^T = \Big[& \overline{C}_{DG}\left(\mathcal{B}_0, x_1^{DG}\right) - \overline{C}_{DG,PR}\left(\mathcal{B}_0, x_1^{DG}\right), \ldots, \overline{C}_{DG}\left(\mathcal{B}_0, x_u^{DG}\right) \\ & - \overline{C}_{DG,PR}\left(\mathcal{B}_0, x_u^{DG}\right), \overline{C}_{DG}\left(\mathcal{B}_0, x_1^{PR}\right) - \overline{C}_{DG,PR}\left(\mathcal{B}_0, x_1^{PR}\right), \ldots, \overline{C}_{PR}\left(\mathcal{B}_0, x_u^{PR}\right) \\ & - \overline{C}_{DG,PR}\left(\mathcal{B}_0, x_u^{PR}\right), 1, 1\Big], \end{aligned} \tag{15}$$

where $\overline{C}_{DG}\left(\mathcal{B}_0, x_i^{\mathrm{DG}}\right)$ denotes the mean covariance between lead content (*Domestic Garden* random function) at location $\mathrm{x}_i^{\mathrm{DG}}$ and locations within block $\mathcal{B}_0$ obtained from the LMCR, $\overline{C}_{PR}\left(x_j^{\mathcal{B}_0,\mathrm{PR}}\right)$ similarly denotes the mean covariance between lead content (*Park* and *Recreational* random function) at location $\mathrm{x}_j^{\mathrm{PR}}$ and locations within block $\mathcal{B}_0$. The term $\overline{C}_{DG,PR}\left(\mathcal{B}_0, x_j^{\mathrm{PR}}\right)$ denotes the mean cross-covariance between location $\mathrm{x}_j^{\mathrm{PR}}$) and block $\mathcal{B}_0$. The mean values of point-to-block covariances were obtained numerically (Webster and Oliver, 2007).

The prediction error variance of the cokriging prediction of the contrast (i.e. the mean square error of the prediction) is

$$\sigma_{\mathrm{K}}^2 = -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} + \overline{C}_{DG}(\mathcal{B}_0, \mathcal{B}_0) + \overline{C}_{DG}(\mathcal{B}_0, \mathcal{B}_0) - 2\overline{C}_{DG,PR},(\mathcal{B}_0, \mathcal{B}_0) \tag{16}$$

where the term $\overline{C}_{DG}(\mathcal{B}_0, \mathcal{B}_0)$ denotes the mean within-block value of the covariance for lead content in the soil of *Domestic Gardens* and

$\overline{C}_{DG}(\mathcal{B}_0, \mathcal{B}_0)$ similarly for *Parks* and *Recreational*. The mean within-block cross-covariance is denoted by $\overline{C}_{DG,PR}(\mathcal{B}_0, \mathcal{B}_0)$. These values were obtained numerically from the respective terms in the LMCR (Webster and Oliver, 2007).

3. Results

Summary statistics for lead content are presented in Table 2. Note that, on the original scales of measurement, both the concentration data and the residuals from the OLS land use class means are strongly positively skewed. The log-transformed data and the residuals from the land use class means for the log-transformed data are symmetrically distributed with small coefficients of skewness and octile skewness values. For this reason subsequent analyses were performed with the log-transformed data.

Table 3 shows cross-validation results for the different variogram estimators for the first iteration of the fitting of the linear mixed model for land use effects, Eq. (1). Iteration 1 is cross-validation of the models fitted to estimates from the OLS residuals. The models fitted to estimates obtained with Dowd's (1984) estimator (shown by the solid discs in Fig. 2) was selected on the basis of the median standardized squared prediction error, and used to obtain a covariance matrix for a refitting of the model by GLS. Iteration 2 is cross-validation of the models fitted to estimates from these GLS residuals, again the model fitted to the estimates obtained with Dowd's (1984) estimator (open circles in Fig. 2) was selected. There was no further change in residuals (the crosses in Fig. 2 are the estimates of the variogram from the residuals of a third iteration), so the variogram model obtained in the second iteration, based on Dowd's (1984) estimator, was selected.

Table 4 shows the model estimates of the land use class means for log lead concentration, with the associated variances. Back-transformation of the class means, under an assumption of log-normality, was done first by simple exponentiation (which gives an estimate of the class median on the original scale) and with the unbiased estimate for the class mean, which, for the kth class is Z_k:

$$\overline{Z}_k = \exp\left\{\overline{Y}_k + \frac{\sigma_k^2}{2}\right\}, \tag{17}$$

where Y_k is the class mean on the log scale and σ_k^2 is the variance of the class mean. Prediction intervals for the class mean on the original scale were obtained by

$$\exp\{\overline{Y}_k \pm (1.96 \times \sigma_k)\}. \tag{18}$$

These results are shown in Fig. 3.

Table 5 presents results for the a priori contrasts set out in section 2.3. Note that the null hypothesis of no difference among the class means can be rejected decisively (contrast 1). This indicates that local land use is an important source of variation in soil lead content. It is also notable that soils in industrial use have significantly larger lead concentrations than do other soils (contrast 2); Table 2 and Fig. 3 show that the mean concentration in the soils under industrial use is the largest of all the class means.

Table 2
Summary statistics of data on lead content of soils of the GLA.

	Pb	log Pb	Pb	log Pb
			Residual from land use class	
	mg kg^{-1}	log$_e$ mg kg^{-1}	mg kg^{-1}	log$_e$ mg kg^{-1}
Mean	296.62	5.26	0	0
Median	184.14	5.22	−75.74	−0.07
Skewness	5.2	0.16	5.42	0.12
Oc Skew	0.51	0.04	0.31	0.1
SD	381.02	0.9	366.91	0.8

Table 3
Median standardized squared prediction error for cross-validation of variogram models fitted to estimates obtained by alternative estimators applied to residuals for log lead concentration from land use means. These are presented for the residuals from the ordinary-least-squares land use means (iteration 1) and then for the residuals from the generalized least squares land-use means (iteration 2), the latter obtained using the selected variogram model (fitted to estimates obtained with the estimator of Dowd) for the OLS residuals.

Median standardized squared prediction error			
Iteration	Variogram estimator		
	Matheron	Cressie–Hawkins	Dowd
1	0.356	0.435	0.465
2	0.352	0.414	0.460

Contrast 3 shows that the null hypothesis of no difference between the model mean concentration in soils under *Parks* and *Recreational* cannot be rejected. However, the lead concentration in the soils of *Domestic Garden* is substantially larger than under *Parks* and *Recreational* considered together. Note that *Domestic Garden* has the second-largest model mean lead content after *Industrial*.

There is no strong evidence against the null hypothesis that *Road Verge* has a larger lead content than does *Urban Open Space*, $P = 0.067$ (contrast 5).

There is strong evidence (contrast 6) that all land in agricultural use has smaller lead concentrations than do the other classes, excluding *Woodland or Forest* and *Industrial*; (contrast 6), and the agricultural land has a significantly larger mean than does *Woodland or Forest* (contrast 7), which (Table 4, Fig. 3) has the smallest mean lead content. There is no evidence for a difference between *Arable* and land under grazing (contrast 8), nor that the lead concentration is different under *Pasture* and *Rough Grazing* (contrast 9). (It is accepted that one factor contributing to the lack of a significant difference between *Pasture* and *Rough Grazing* may be the difficulty of distinguishing these two land uses consistently in the field.)

Table 6 shows results for cross-validation of the auto-variogram models fitted to estimates from different estimators for the lead content in classes *Domestic Garden* and *Park* or *Recreational*. In both cases the

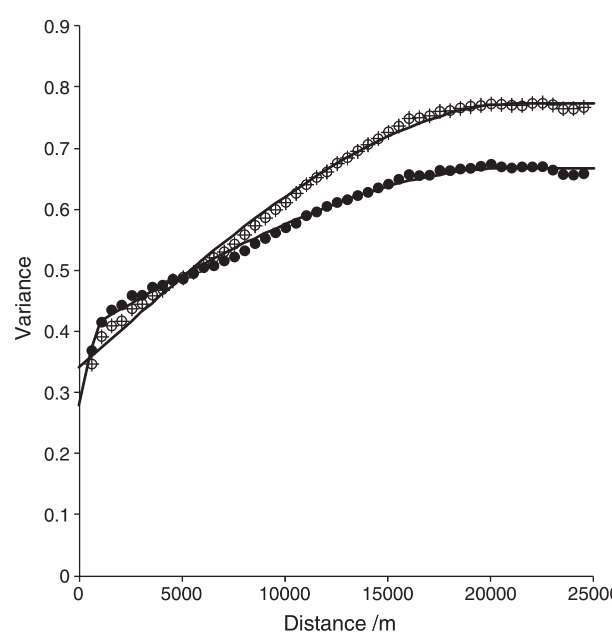

Fig. 2. Empirical variograms for variation of lead content within land use classes. The solid discs are for the OLS residuals, the open circles for the second iteration (GLS residuals) and the cross for the third iteration. Fitted models are shown for the OLS residuals and the third iteration.

Table 4
Model estimates of land use class mean log lead concentration, with associated variances.

Land use	Mean mg kg^{-1}	Variance $\log_e$ mg kg^{-1}
Arable	4.63	0.041
Commercial and residential	5.08	0.042
Domestic garden	5.28	0.040
Industrial	5.42	0.046
Other	5.08	0.041
Park	4.73	0.040
Pasture	4.69	0.042
Recreational	4.79	0.040
Road verge	5.15	0.040
Rough grazing	4.73	0.040
Urban open space	5.09	0.040
Woodland or forest	4.46	0.041

results for the SRD estimator gave results closest to the expected value for a sound variogram model with normal kriging errors. This estimator was therefore selected, and the fitted linear model of coregionalization is shown in Fig. 4.

Fig. 5 shows the cokriged estimates of lead concentration for soils in the two land use classes considered. Note that in all cases these are the best linear unbiased predictions of lead concentration at a location conditional on the land use at that location, which is not known everywhere but only at sample sites. The maps in Figs. 5–7 therefore show continuous variation, not discrete step changes. These maps show, particularly for *Domestic Gardens* a marked increase in lead concentration in central parts of the GLA relative to the margins. Fig. 6 shows the corresponding estimates of the difference between the lead concentrations (log-scale) in *Domestic Gardens* and *Park* and *Recreational* considered together. Note that the difference is positive (*Domestic Garden* is larger) almost everywhere, but that the contrast is largest over a region in south-west London, and also some smaller parts of the north and east. To indicate the uncertainty of the local predictions of this contrast,

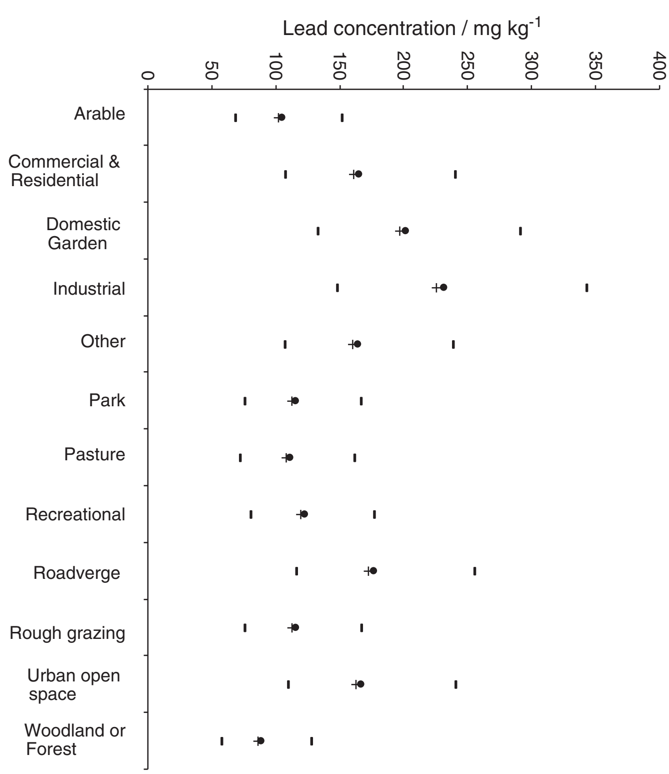

Fig. 3. Lead concentration by land use class in the original scales of measurement. The horizontal bars are 95% confidence intervals for the model mean. The crosses are estimates of the model median and the solid disks unbiased estimates of the model mean.

Table 5
Results for tests on contrasts between land uses with respect to mean log lead concentration in the soil.

	Contrast	Wald statistic	P^a
1	Difference among all classes	880.5	≪0.001
2	*Industrial*[b] vs all other classes	43.8	≪0.001
3	*Parks* vs *Recreational*	2.3	0.131
4	*Parks* or *Recreational* vs *Domestic Garden*[b]	508.1	≪0.001
5	*Road verge* vs *Urban open space*	3.4	0.067
6	*Arable* or *Pasture* or *Rough Grazing* vs other classes[b] (excluding *Woodland or Forest* and *Industrial*	106.0	≪0.001
7	*Arable*[b] or *Pasture*[b] or *Rough Grazing*[b] vs *Woodland or Forest*	22.6	≪0.001
8	*Arable* vs *Pasture* or *Rough Grazing*	2.2	0.139
9	*Pasture* vs *Rough Grazing*	0.44	0.508

[a] *P*-value for the null hypothesis that the compared means are equal.
[b] Class(es) with the larger mean concentration where there is a significant difference.

Table 6
Median standardized squared prediction error for cross-validation of auto-variogram models for log lead concentration in soils of land use classes *Domestic Garden* and *Park* or *Recreational.*

Median standardized squared prediction error			
Land use	Estimator		
	MoM	SRD	MAD
Domestic Gardens	0.387	0.443	0.495
Park or *Recreational*	0.355	0.482	0.588

Fig. 7 shows the predictions standardized by the kriging standard error — the square root of the kriging variance in Eq. (16). The scale shows, in effect, where the 95% confidence interval of the prediction excludes zero (smallest and largest of the four intervals) and where the prediction is negative but the limits include zero, and where it is positive, but includes zero. Note that this Figure should not be interpreted as showing an overall region where the mean contrast is significantly larger than zero, it shows, rather, point confidence intervals.

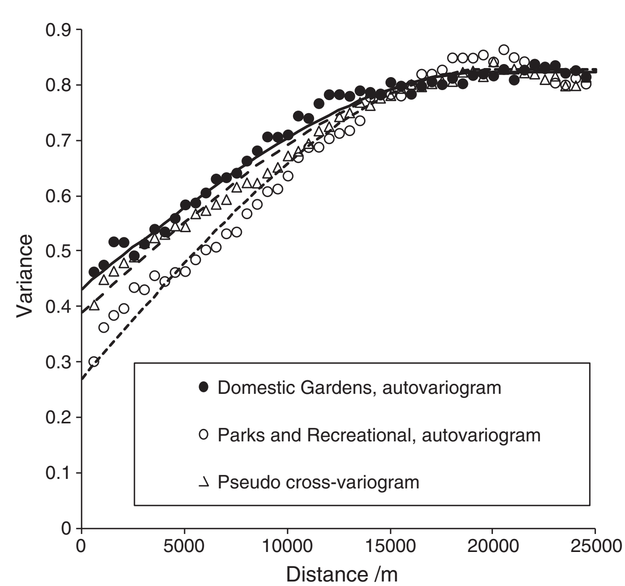

Fig. 4. Empirical estimates of the auto-variogram and pseudo cross-variogram (by the SRD estimator) for lead content of the soils in land use classes *Domestic Garden* and *Park* or *Recreational* considered together. The solid and dashed lines show the fitted linear model of coregionalization.

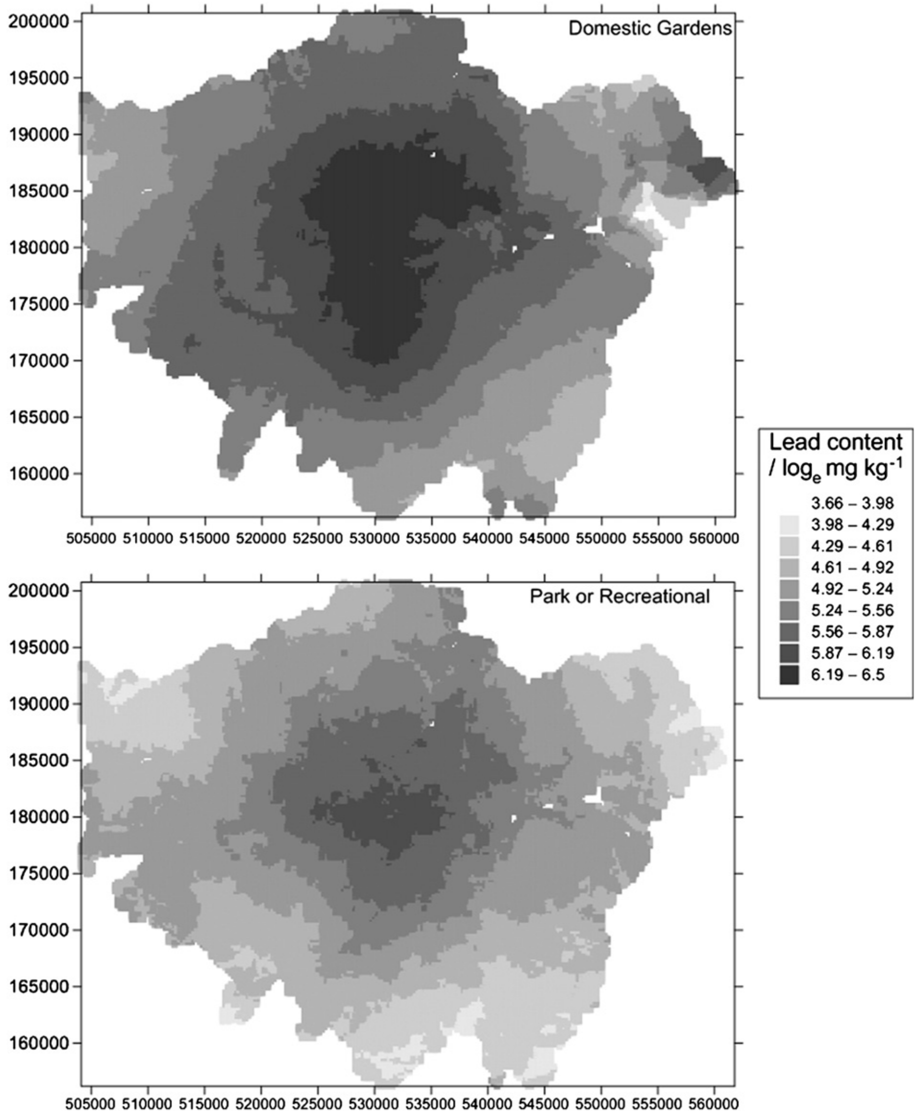

Fig. 5. Cokriged estimates of lead concentration (log scale) under land use classes *Domestic Garden* and *Park* or *Recreational* considered together.

4. Discussion

It is not surprising that land use is an important source of variation in the lead concentration in the soil of London. However, the overall test was worth performing for two reasons. First, the recorded land use is what the sampling teams observed and recorded. It takes no account of past land use, which may be complex in the urban environment (Rawlins et al., 2005). Land use effects may be obscured, for example, where previous open space has been developed, or, conversely, former industrial sites have been converted to recreational use. Second, it is known (McGrath and Loveland, 1992) that atmospheric deposition of lead has been very important, particularly historically before the ban on lead additives to fuel. Since this is a diffuse process it might be expected to obscure local land use effects to some degree.

Although few sampled sites are under industrial land use (Table 1), these results (contrast 2) show that local pollution as a result of industrial activity can lead to large lead concentrations in urban soils. It is also notable that lead concentration in domestic gardens is large, on average. This is consistent with previous studies. Thornton et al. (1990) report a survey on lead content in soils of residential areas across the UK. Lead concentration in domestic garden soils was consistently larger than in gardened soils in public parks, and this elevation was largest for old dwellings. The reasons why this may be expected were noted in section 2.3. The first source is the burning of domestic waste and spreading of ash (Thornton, 1991). The second is the use of lead-based paints which can be a very significant source of soil contamination when old paintwork is sanded or scraped (Mielke et al., 2001). These factors are now historical: the burning of domestic waste and spreading of ash is no longer widely practiced in cities due to universal provision of waste collection by local authorities and bans on bonfires and incineration in smoke-free zones. Similarly, lead compounds have been banned in paint in the United Kingdom since 1992 — Environmental Protection (Controls on Injurious Substances) Regulations (1992). None the less, it remains important for local authorities, policy makers and others to be aware of this legacy of pollution. In particular, given the recent emphasis on the ecosystem services delivered by urban green spaces (Davies, 2011) it is relevant to note (contrast 4) that the quality of soils in parks and recreational areas is, in general, higher (at least with respect to lead pollution) than that of domestic gardens across the GLA. Similarly we note that, despite the importance of deposition of atmospheric lead — which may explain why roadside verges are not significantly enhanced relative to urban open space (contrast 5) —

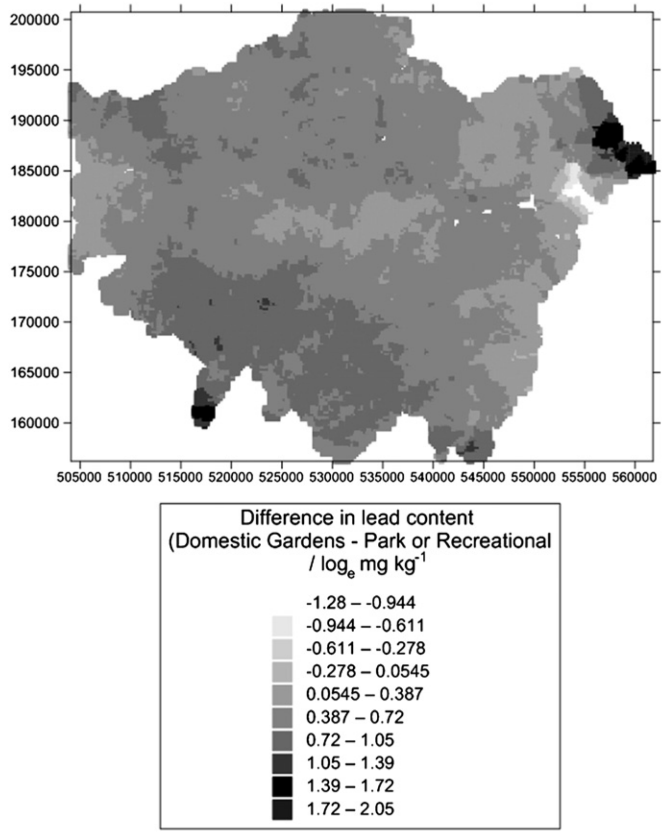

Fig. 6. Cokriged estimate of the difference between log lead concentration under land use classes: *Domestic Garden–Park* or *Recreational* considered together.

agricultural land within the GLA contains significantly less lead than other land uses (contrast 6), and woodland and forest in the GLA (contrast 7) contains significantly less lead again than the agricultural land.

While various studies on urban soils have presented concentrations of lead (Johnson et al., 2011), few have done so on the basis of a land use classification, and where they do it is not always clear that the land use classes are directly comparable to those used in the London Earth survey. There are two studies which, with this latter reservation, it is interesting to compare with the results in this paper. In a survey of the soils of Dublin (Republic of Ireland) Glennon et al. (2012) determined lead concentrations by inductively coupled plasma atomic emission spectroscopy (ICP-AES). These results are not directly comparable with the XRF determinations reported here. Median concentrations in topsoil for land under Heavy Industry, Residential use, Managed Open Land and Unmanaged Open Land were, respectively 174 mg kg $^{-1}$, 70 mg kg $^{-1}$, 80 mg kg $^{-1}$ and 50 mg kg $^{-1}$. There are all smaller concentrations than for comparable land uses in the GLA, and it is interesting to note that the concentration for Residential land falls between that for the two Open Land categories. Birke et al. (2011) report results from a survey in Berlin (Germany) where concentrations were determined by XRF. Median concentrations for Industrial sites, Dense Residential sites, Agricultural land and Woodland were 87 mg kg $^{-1}$, 109 mg kg $^{-1}$, 25 mg kg $^{-1}$, and 26 mg kg $^{-1}$ respectively. Again, these are smaller than comparable values for the GLA, but it is interesting to note the large concentrations associated with dense residential areas.

These inferences are about the overall model means for soils in the GLA. The geostatistical investigation described in Section 2.4, shows that there may be substantial spatial variation about the overall mean. In particular, for both groups of land uses investigated, the lead content of the soils of central London are clearly larger than the content of soils in outer London. The contrast between *Domestic Garden* and *Park* and *Recreational* considered together is also spatially variable. The difference is largest over a region of South-West London, as well as some more isolated regions in the North and North-East. It is notable that the South-West region includes two large Royal Parks (Bushy Park and Richmond Park) as well as Wimbledon Common. It is likely that the contrast with *Domestic Garden* is particularly large here because of the large area of these parks, which means that much of the soil is at a greater distance from the road network and other sources than would be the case for soils in smaller parks and recreational grounds elsewhere. There is a large Royal Park in central London (Hyde Park, with British National Grid coordinates close to 525000, 180000) but the contrast with Domestic Gardens is not large here, perhaps reflecting the fact that all soils in central London are or have been subject to large deposition rates of lead regardless of land use. This information on the spatial variation of contrasts may be relevant to local authorities and policy makers. It highlights that land use effects are not uniform, and that the quality of soils in Parkland and other urban green space may reflect other spatially variable factors which may interact with land use.

This study illustrates how appropriate spatial statistical analysis allows sound inferences to be drawn from data in an urban geochemical survey. The original data were collected by non-probability sampling, so a conventional analysis of variance, with assumptions of independence, would not be valid. However, a model-based analysis is possible, as illustrated here. Furthermore, the systematic sampling design facilitates the geostatistical mapping of variables such as the land use contrast investigated here. This methodology may be useful for addressing other questions about the soils of London and other urban centres with similar data.

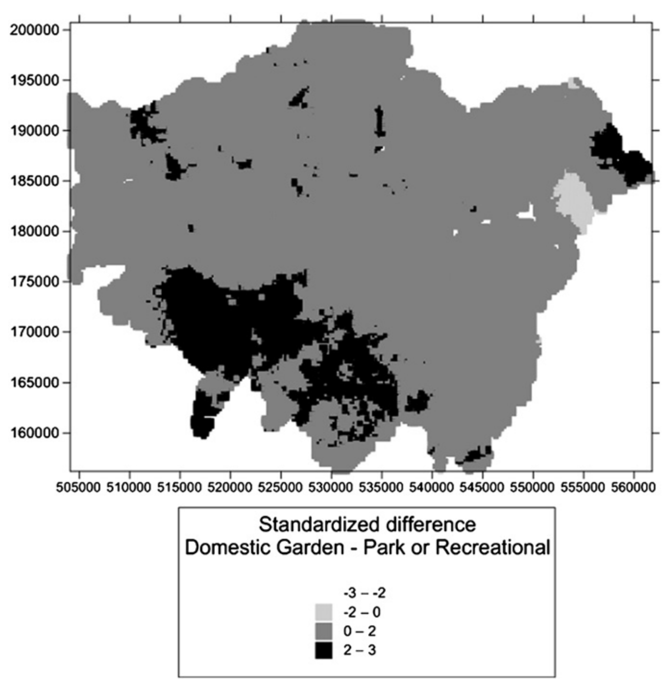

Fig. 7. Cokriged estimate of the difference between log lead concentration under land use classes standardized by the kriging standard error.

5. Conclusions

The soils of London show variation in their lead content. Land use, as recorded at the time of sampling, is a significant source of variation in this variable, and notable effects are the relative enrichment of sites in industrial use and domestic gardens. Agricultural soils in the GLA have smaller lead content than most other land uses, but more lead than woodland or forest soils which have the smallest overall mean. Land use effects are spatially variable, and the contrast between domestic gardens, and soils of parks and recreational land, while significant overall, is larger in some parts of London than elsewhere. This study has shown how a model-based analysis of a structured set of hypotheses permits sound inference and insight into factors controlling quality of urban soils.

Acknowledgments

The contribution of all British Geological Survey staff and student volunteers involved in the geochemical sampling of London, and the subsequent soil analysis, is acknowledged. We are grateful to Dr Chris Johnson for comments on the manuscript and to Dr Steve McGrath for helpful discussions. This paper is published with the permission of the Director of the British Geological Survey.

References

Bain, D.J., Hale, R.L., Wollheim, W.M., 2012. Hotbeds of biogeochemical diversity: insights from urban long-term ecological research sites, 8, pp. 435–438.
Birke, M., Rauch, U., Stummeyer, J., 2011. Urban geochemistry of Berlin, Germany. In: Johnson, C.C., Demetriades, A., Locutura, J., Ottesen, R.T. (Eds.), Mapping the Chemical Environment of Urban Areas. Wiley-Blackwell, Chichester, pp. 245–268.
Bishop, T.F.A., Lark, R.M., 2007. A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season. Geoderma 141, 384–396.
Brys, G., Hubert, M., Struyf, A., 2003. In: Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P.J. (Eds.), A comparison of some new measures of skewness. In: Developments in Robust Statistics. Physica-Verlag, Heidelberg, pp. 98–113.
Bullock, P., Gregory, P.J. (Eds.), 1991. Soils in the Urban Environment. Blackwell, Oxford.
Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. Journal of Soil Science 31, 333–341.
Chen, X., Xia, X.H., Zhao, Y., Zhang, P., 2010. Heavy metal concentrations in roadside soils and correlation with urban traffic in Beijing, China. Journal of Hazardous Materials 181, 640–646.
Clark, H.F., Brabander, D.J., Erdil, R.M., 2006. Sources, sinks, and exposure pathways of lead in urban garden soil. Journal of Environmental Quality 35, 2066–2074.
Craul, P.J., 1985. A description of urban soils and their desired characteristics. Journal of Arboriculture 11, 330–339.
Cressie, N., Hawkins, D., 1980. Robust estimation of the variogram. Journal of the International Association for Mathematical Geology 12, 115–125.
Davies, L., 2011. Urban. Chapter 10 of UK National Ecosystem Assessment, 2011. Technical Report. UNEP-WCMC, Cambridge.
Davies, B.E., Conway, D., Holt, S., 1979. Lead pollution of London soils: a potential restriction on their use for growing vegetables. Journal of Agricultural Science 93, 749–752.
de Gruijter, J.J., Brus, D.J., Biekens, M.F.P., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer, Berlin.
Deutsch, C.V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York.
Dobson, A.J., 1990. An Introduction to Generalized Linear Models. Chapman & Hall, London.
Dowd, P.A., 1984. The variogram and kriging: robust and resistant estimators. In: Verly, G., David, M., Journel, A.G., Marechal, A. (Eds.), Geostatistics for Natural Resources Characterization. D. Reidel, Dordrecht, pp. 91–106 (Part 1).
Environmental Protection (Controls on Injurious Substances) Regulations, 1992. Statutory Instrument 1992 No. 31. (United Kingdom, available at) http://www.legislation.gov.uk/uksi/1992/31/made.
Fordyce, F.M., Brown, S.E., Ander, E.L., Rawlins, B.G., O'Donnell, K.E., Lister, T.R., Breward, N., Johnson, C.C., 2005. GSUE: urban geochemical mapping in Great Britain. Geochemistry: Exploration, Environment, Analysis 5, 325–336.
Glennon, M., Scanlon, R.P., O'Connor, P.J., Finne, T.E., Andersson, M., Eggen, O., Jensen, H.K.B., Ottesen, R.T., 2012. Geochemical baseline for heavy metals and organic pollutants in topsoils in the greater Dublin area. Technical Report Geological Survey of Ireland.Beggars Bush, Dublin.
Guney, M., Zagury, G.J., Dogan, N., Onay, T.T., 2010. Exposure assessment and risk characterization from trace elements following soil ingestion by children exposed to playgrounds, parks and picnic areas. Journal of Hazardous Materials 182, 656–664.
Johnson, C.C., 2005. 2005 G-BASE Field Procedures Manual. British Geological Survey Internal Report OR/05/097. British Geological Survey, Keyworth.
Johnson, C.C., Demetriades, A., Locutura, J., Ottesen, R.T., 2011. Mapping the Chemical Environment of Urban Areas. Wiley-Blackwell, Chichester.
Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.
Knights, K.V., Scheib, A.J., 2010. London Earth: details of field campaigns across the Greater London area, 2005 to 2009. British Geological Survey Internal Report OR/09/056. British Geological Survey, Keyworth.
Laidlaw, M.A.S., Filippelli, G.M., 2008. Resuspension of urban soils as a persistent source of lead poisoning in children: a review and new directions. Applied Geochemistry 23, 2021–2039.
Lark, R.M., 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.
Lark, R.M., 2002. Robust estimation of the pseudo cross-variogram for cokriging soil properties. European Journal of Soil Science 53, 253–270.
Lark, R.M., Cullis, B.R., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55, 799–813.
Lark, R.M., Bellamy, P.H., Rawlins, B.G., 2006. Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. Geoderma 133, 363–379.
Marchant, B.P., Tye, A.M., Rawlins, B.G., 2011. The assessment of point-source and diffuse soil metal pollution using robust geostatistical methods: a case study in Swansea (Wales, UK). European Journal of Soil Science 62, 346–358.
Matheron, G., 1962. Traité de Géostatistique Appliqué, Tome 1. Mémoires du Bureau de Recherches Géologiques et. Minières, Paris.
Maurice, P.A., 2009. Environmental Surfaces and Interfaces from the Nanoscale to the Global Scale. John Wiley & Sons, Hoboken N.J.
McGrath, S.P., Loveland, P.J., 1992. The Soil Geochemical Atlas of England and Wales. Blackie Academic and Professional, London.
Mielke, H.W., Powell, E.T., Shah, A., Gonzales, C.R., Mielke, P.W., 2001. Multiple metal contamination from house paints: consequences of power sanding and paint scraping in New Orleans. Environmental Health Perspectives 109, 973–978.
Mielke, H.W., Gonzales, C.R., Powell, E.T., Jartun, M., Mielke, P.W., 2007. Nonlinear association between soil lead and blood lead of children in metropolitan New Orleans, Louisiana: 2000–2005. Science of the Total Environment 388, 43–53.
Miranda, M.L., Kim, D., Galeano, M.A.O., Paul, C.J., Hull, A.P., Morgan, P., 2007. The relationship between early childhood blood lead levels and the performance on end-of-grade tests. Environmental Health Perspectives 115, 1242–1247.
Myers, D.E., 1983. Estimation of linear combinations and cokriging. Mathematical Geology 15, 633–637.
Papritz, A., Flühler, H., 1994. Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 62, 29–43.
Papritz, A., Künsch, H.R., Webster, R., 1993. On the pseudo cross variogram. Mathematical Geology 25, 1015–1026.
Payne, R.W. (Ed.), 2010. GenStat Release 13 Reference Manual Part 3: Procedure Library PL21. VSN International, Hemel Hempstead, Hertfordshire.
Rawlins, B.G., Lark, R.M., O'Donnell, K.E., Tye, A., Lister, T.R., 2005. The assessment of point and diffuse soil pollution from an urban geochemical survey of Sheffield, England. Soil Use and Management 21, 353–362.
Rawlins, B.G., McGrath, S.P., Scheib, A.J., Breward, N., Cave, M., Lister, T.R., Ingham, M., Gowing, C., Carter, S., 2012. The Advanced Soil Geochemical Atlas of England and Wales. British Geological Survey, Keyworth, Nottingham.
Thornton, I., 1991. Metal contamination of soil in urban areas. In: Bullock, P., Gregory, P.J. (Eds.), Soils in the Urban Environment. Blackwell, Oxford, pp. 47–75.
Thornton, I., Davies, D.J.A., Watt, J.M., Quinn, M.J., 1990. Lead exposure in young children from dust and soil in the United Kingdom. Environmental Health Perspectives 89, 55–60.
UK National Ecosystem Assessment, 2011. Technical Report. UNEP-WCMC, Cambridge.
Webster, R., McBratney, A.B., 1989. On the Akaike information criterion for choosing models for variograms of soil properties. Journal of Soil Science 40, 493–496.
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd Edition. John Wiley & Sons, Chichester.